Electrophysiological and neuroendocrine correlates of trust in the Investment Game

Adrian R. Willoughby[1], Jorge A. Barraza[2], Harold Javitz[1], Brian J. Roach[3,4], M. Travis Harrison[1], Massimiliano de Zambotti[1], James C. Cox[5], John Murray[1], Paul J. Zak[2], Judith M. Ford[3,4], Ian M. Colrain[1], Gregory Myers[1]

[1]SRI International, Menlo Park, CA, [2]Center for Neuroeconomics Studies, Claremont Graduate University, Claremont, CA, [3]Psychiatry Service, San Francisco VA Medical Center, San Francisco, CA, [4]Department of Psychiatry, University of California, San Francisco, CA, [5]Experimental Economics Center, Georgia State University, Atlanta, GA

Correspondence to: adrian.willoughby@sri.com

Abstract

In this study, we made an extensive analysis of electrophysiological and neuroendocrine correlates of interpersonal trust. Prior to playing the Investment Game—an economic decision-making game in which participants can entrust some of their money to a partner with the hope of future reciprocity—the two participants had the opportunity to discuss how much they would send to each other and how they would reciprocate. We recorded EEG, ECG, respiration, and skin conductance, and measured blood oxytocin, adrenocorticotrophic hormone (ACTH), and cortisol levels immediately following this interaction in order to investigate how these physiological measures were associated with trust behavior in the subsequent decision-making game (i.e., how much the participant sent to her partner). The results showed a few, weak relationships: most notably, reduced skin-conductance level (SCL) and increased ACTH were associated with increased trusting behavior. We tentatively propose that SCL may co-vary with trust whereas ACTH may be sensitive to risk.

Introduction

Interpersonal trust is a complex psychosocial construct that has been subject to sociological, social psychological, economic, and organizational research (see Lewicki, Tomlinson, & Gillespie, 2006, for a review of models of trust). In recent years, it has also become an area of increasing interest for cognitive neuroscientists investigating the cognitive and brain processes involved in trust. The definition of trust has evolved as research has progressed, and most researchers now agree on its main components. A reasonable basic definition (adapted from Mayer, Davis, & Schoorman, 1995) is that trust is "the willingness to accept risk based on the expectations of another person's future behavior." This definition captures several important facets of trust: it requires risk on behalf of the trustor, the trustor has no control over the actions of the trustee, and the trustor has an expectation that the trustee will behave in a particular way in the future.

Since interpersonal trust requires two participants, characteristics of the individuals involved and the nature of their relationship affect the degree of trust between them. The trustor's personality is one important factor; research has shown that the general personality factor 'agreeableness' and the narrower personality trait 'propensity to trust' both affect trusting behavior (Mooradian, Renzl, & Matzler, 2006). Important characteristics of the trustee can be broadly divided into two categories: cognitive and affective (McAllister, 1995). The cognitive category encompasses characteristics such as the trustee's perceived competence and ability to execute the expected behavior successfully. The affective category refers to characteristics such as the perceived benevolence that could motivate the trustee to actually perform the entrusted task. Finally, the relationship between the participants is also an important factor. If participants have a long history of trust and reciprocity, they will naturally trust each other more than two participants with little or no prior history.

Because trust is a multifaceted concept, measuring it accurately can be difficult. One technique that has been used to measure trust and its constituent components is attitudinal questionnaires (Evans & Revelle, 2008). While this technique has the advantage that it can measure the different components of trust, it also suffers from the problems of questionnaire methods in general; including social desirability effects and consistent findings that people's beliefs (as expressed in a questionnaire) don't predict their actual behavior. Given these limitations, several experimental methodologies, particularly in the field of behavioral economics, have been developed to quantify trust and investigate how the various characteristics of the trustor, the trustee, and their relationship affect it. One particular methodology uses interpersonal economic decision-making tasks, the most popular of which (and the one used to assess trust in this study) is the Investment Game (IG) (Berg, Dickhaut, & McCabe, 1995). In this game, experimenters endow two participants with an equal amount of money; then one participant (the First Mover—FM) has the opportunity to transfer none, some, or all of her money to the other participant (the Second Mover—SM). During the transfer, the amount of money is trebled. Now the SM has the opportunity to return some money back to the FM. This game captures the critical features of trust: the risk assumed by the FM, her subsequent lack of control, and her expectation of the SM's future behavior. The control condition for the IG is typically the Dictator Game (DG), in which only one participant in the dyad makes a decision: how much to send the other participant (which is again trebled). This matches the FM's decision in the IG; however, in the DG the participant sending money has no expectation of getting anything back from her partner. This task accounts for participants' general altruism, so that any additional money sent in the IG can be attributed to trust—if the FM doesn't trust the SM to return any money, she will not send any amount beyond that motivated by altruism; if she trusts them completely, the FM will send all of her money to the SM. While behavioral measures of trust, such as the IG, overcome the limitations of questionnaire approaches to measuring trust, they are not without their own problems. They are only a summary measure of trust and do not interrogate the various constituent components of trust. Furthermore, while questionnaires are subject to social desirability biases, behavioral measures of trust are also subject to other forces, such as social norms and pressures. However, research has shown that behavioral measures of trust are correlated with well-designed attitudinal measures and they are relatively insensitive to social desirability effects (Evans & Revelle, 2008; Naef & Schupp, 2009).

While behavioral economics research using this measure is common, researchers have only recently started to use it in conjunction with neurophysiological methods. To date, most neuroscience research into trust has used functional neuroimaging and neuroendocrinology—the latter usually focused on oxytocin (OT). fMRI research has demonstrated that a number of brain regions are active during financial decision-making games. The ventral striatum appears to activate when participants display trust and experience reciprocation, whereas the insula is active in response to inequity and non-reciprocity (Sanfey, 2007). Researchers have reported an increase in amygdala activity when participants judge someone untrustworthy, whereas the superior temporal sulcus is more active when they rate the person as trustworthy (Winston, Strange, O'Doherty, & Dolan, 2002). In one study focusing on trust decisions, Krueger et al. (2007) observed increased activity in the paracingulate cortex (PcC) and the septal area during trust decisions. They proposed that the activity in the PcC was important in attributing mental states to other people (i.e., theory of mind). The septal area is connected with the hippocampus and ventral tegmental area, suggesting it plays a role in combining reward information with its context (Luo, Tahsili-Fahadan, Wise, Lupica, & Aston-Jones, 2011), and is also involved in the regulation of OT release (Lee, Macbeth, Pagani, & Young, 2009; Powell & Rorie, 1967). These results suggest that trust involves a number of brain circuits related to the processing of emotion, reward, and theory of mind.

OT is a neuropeptide hormone that researchers originally found to be associated with maternal behavior and pair bonding (Insel, Winslow, Wang, & Young, 1998). However, more recent research has provided evidence for a role in trust and mutual cooperation (Zak, Kurzban, & Matzner, 2004, 2005). Kosfeld, Heinrichs, Zak, Fischbacher, and Fehr (2005) demonstrated that intranasal administration of OT increased trust displayed by the FM in the IG. Importantly, OT affected neither risk-taking in a control task nor the SM's reciprocity behavior, suggesting that OT modulated trust behavior in particular, not just risk-taking or altruism in general. In a combined OT and fMRI study, Baumgartner, Heinrichs, Vonlanthen, Fischbacher, and Fehr (2008) showed that OT administration decreased activity in the caudate, amygdala, and midbrain regions following feedback that showed moderately non-reciprocal behavior in a trust game, but only when responses were attributed to another person, not a computer. These regions of OT modulation are consistent with parts of the reward and emotion circuits active during trust in neuroimaging research. A meta-analysis of 23 studies investigating OT administration and trust has shown that OT administration results in enhanced recognition of facial emotions and an increase in in-group trust, but no effect on out-group trust (van IJzendoorn & Bakermans-Kranenburg, 2012). These results suggest that OT may exert its effect through altering perception of a partner's trustworthiness and/or reducing fear of betrayal.

The theoretical analysis and experimental literature implicate both cognitive and emotional processes in interpersonal trust. Cognitive processes include creating a mental model of the trustee and deciding to what extent she is capable and willing to perform the desired action. The emotional processes include those related to assessing risk and imagining how you might feel if your trust is betrayed, or your emotional reaction to the trustee in general. Existing research investigating emotional responses in decision-making under different risk conditions, which is a major factor in behavioral measures of trust, suggest different neural structures are involved in risk-taking in different contexts (Levin et al., 2012).

This study aimed to make an extensive, although exploratory, assessment of electrophysiological and neuroendocrine responses associated with trust, using the IG to measure trust behavior. One major difference in this research from that conducted previously is the time at which the physiological measures are taken. Much previous research has focused on either the choice or outcome phase of the decision-making task (Schilke, Reimann, & Cook, 2013). While physiological measurements at the choice phase are close to the trusting behavior, they may not be close to the actual decision to trust. It is possible that this decision may have been made somewhat earlier than the time at which the behavior was performed (e.g., on an earlier meeting between the two participants, or on an earlier trial when trust had been validated or betrayed). In the experiment reported here, we made the physiological measurements directly after the dyadic interaction (somewhat prior to the IG) with the hope of capturing the physiological effects of those processes surrounding the trust decision itself. Of course, it is difficult to know exactly when the internal decision to trust is made, or indeed if such a single process exists in a complex psychological process like trust.

Because previous research has indicated there is a strong association between OT and trust (Zak et al., 2004, 2005), OT was our primary hormonal measure. In addition, however, we also measured cortisol and adrenocorticotrophic hormone (ACTH), both of which are associated with the stress response. Based on previous research, we expected to see increased levels of OT associated with trust (Zak et al., 2004, 2005) and, because OT plays a role in reducing stress (Heinrichs, Baumgartner, Kirschbaum, & Ehlert, 2003), decreased levels of cortisol and ACTH.

Neural circuits responsible for emotional processing are also associated with trust; therefore, we recorded peripheral physiological measures that reflect sympathetic and parasympathetic system activity: heart rate (HR), heart rate variability (HRV), and electrodermal activity (EDA). Because increased HR and EDA are associated with emotional arousal, we expected them to decrease with trust. HRV measures beat-to-beat changes in HR and decreases with stress and anxiety; therefore, we predicted that trust would be associated with an increase in HRV.

One important behavioral feature associated with an emotional response is whether to approach or withdraw from the stimulus responsible for generating it. That is, seeing something that evokes a negative emotion typically elicits an avoidance response, whereas seeing a stimulus that evokes a positive emotion elicits an approach response. An electroencephalographic (EEG) phenomenon known as frontal alpha asymmetry is related to approach and avoidance behaviors (Harmon-Jones & Allen, 1998). Alpha activity is the 8–12Hz frequency band of the EEG spectrum, and relatively greater alpha power at left frontal electrode sites is associated with approach behavior, whereas relatively greater power over right frontal sites is associated with avoidance behavior. Therefore, we predicted a relative increase in left frontal alpha power associated with trust, reflecting the approach behavior associated with a positive emotional state.

Methods

Procedure

Four participants completed two four-hour experimental sessions one week apart. Following intravenous (IV) catheter fitting and electrophysiological instrumentation, participants were awarded a financial endowment of $120 that they used during the subsequent economic games. Participants played both the IG and DG (described in detail later) with two other participants (i.e., four games in total). Before playing the IG, participants engaged in a face-to-face interaction with their partner during which we encouraged them to discuss strategy and make a promise about how much money they would send each other during the game. Immediately following the interaction period, we drew blood and recorded neurophysiological measures during an eight-minute quiet period (four minutes of which the participant spent with their eyes open and four minutes with their eyes closed—conditions typically used when measuring resting-state EEG power). At the end of the session, one of the games was randomly chosen to determine participant payment. We counterbalanced the order of the economic games, first and second mover order, and partner order across experimental groups and sessions. Figure 1 shows the full procedure for one session in one counterbalancing group.

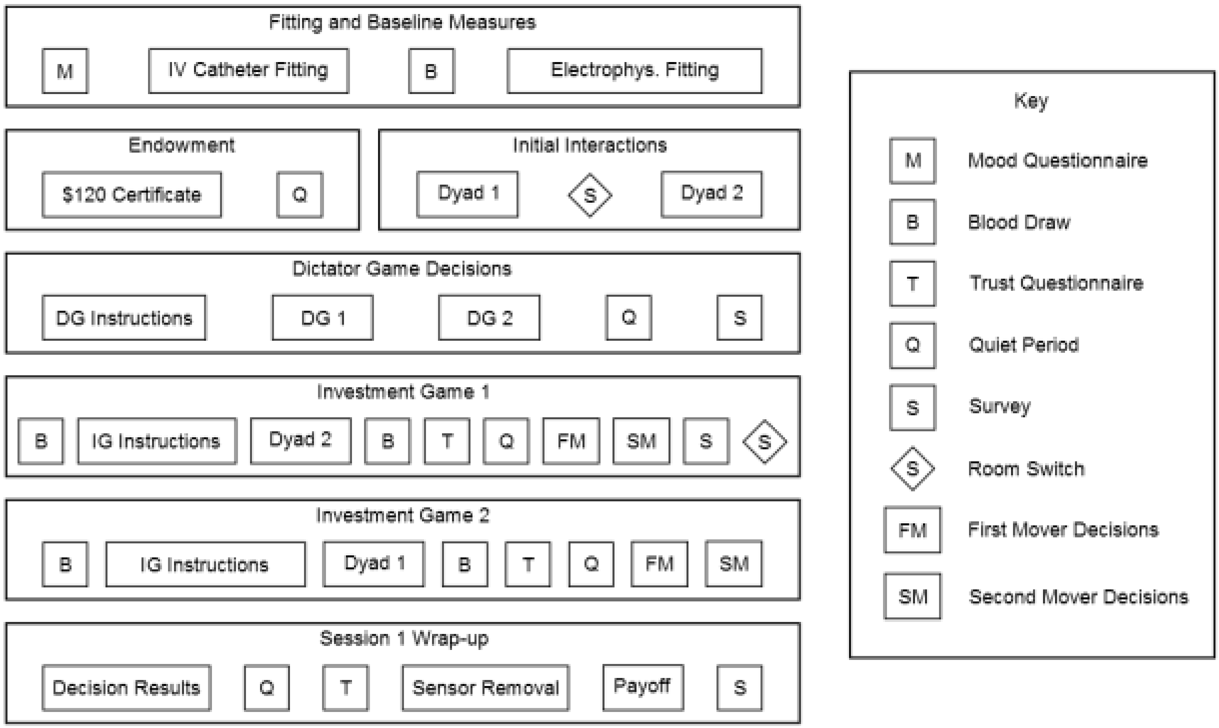


Figure 1. Full experimental procedure for Protocol A, Session 1 of one counterbalancing group. Protocol B used a short introduction instead of the initial face-to-face interactions. Session 2 included no initial interactions.

We developed two variants of the experimental protocol. In Protocol A, we recruited participants as pairs of friends; two pairs who did not know each other participated as a group of four. In each session, each participant interacted with both her friend and a stranger from the other pair. In Protocol B, we recruited participants individually so all were strangers; each participant interacted with two of the other participants in the group of four. Another difference between the protocols was that in Protocol A participants had a five-minute, unstructured interaction period prior to performing any tasks, which was reduced to a brief group introduction in Protocol B.

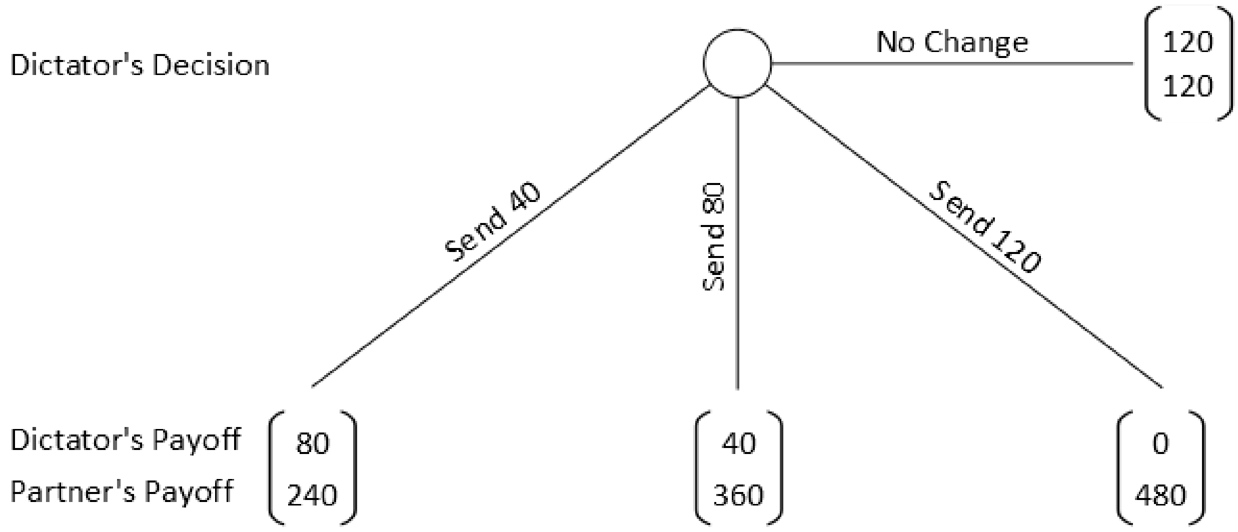
Participants

In Protocol A, 52 participants completed Session 1, and 48 (30 males, mean age: 31.9 years; 18 females, mean age: 33.6) completed both sessions (i.e., only one Session 2 was not completed). In Protocol B, 100 (54 males, mean age: 31.0; 46 females, mean age: 33.3) out of 116 participants completed both sessions; 54 participants were born and raised in the USA and 46 were born and raised outside the USA. Participants were considered raised abroad if they moved to the USA after they turned 16. All participants were aged 25-50, were fluent in English, and gave their informed consent prior to participating.

Behavioral measures

Participants completed two financial decision-making games, the Dictator Game (DG) and the Investment Game (IG). In the DG, participants simply decided how much of their $120 endowment to give to their partner without an expectation of receiving anything back. The experimenters trebled this amount before paying the partner. In order to reduce variability, the experimental setup only allowed four decision options: participants could send $0, $40, $80 or $120 of their endowment to their partner (resulting in $0, $120, $240, or $360 paid to their partner).

The IG requires both players to make a decision. The FM makes the same decision as in the DG—that is, how much of her endowment to send to the SM. Again, the amount sent is trebled. Unlike in the DG, the SM now has an opportunity to return some of the money they received to the FM. To limit decision variability, we allowed only three SM options: they could keep all the money, return the amount that the FM sent, or split the total equally. Figure 2 shows the options available to the players in the DG (upper panel) and IG (lower panel). Participants played as both the FM and SM for both dyadic partnerships, and made SM decisions for each possible FM choice of their partner. Participants made their decisions for each game simultaneously, so that the SM didn't know what decision the FM had made. Therefore, they were asked to make their SM choices for all potential FM decisions (e.g., 'If your partner sends you $40/$80/$120 as FM, what will your SM decision be?'). At the end of the session, participants learned their partners' decisions for only one game (out of the four) randomly chosen to determine payouts.



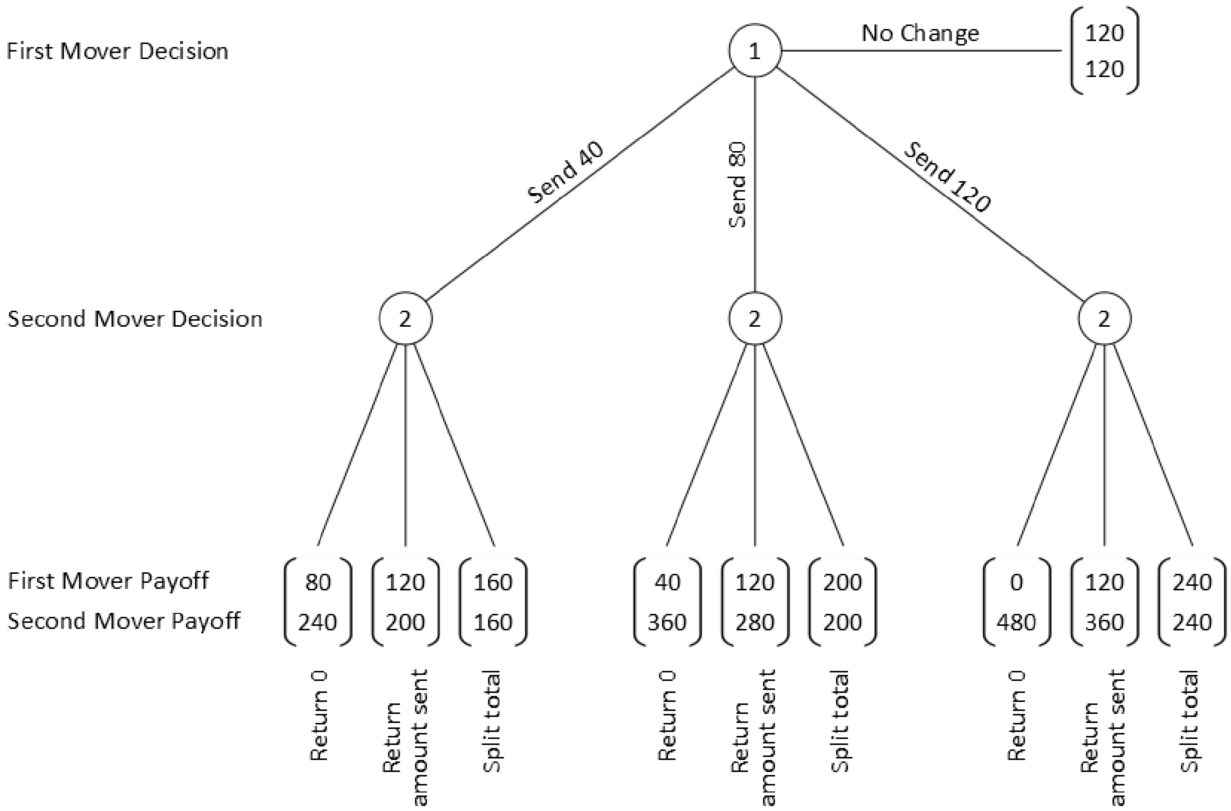


Figure 2. The upper panel shows the choices available in the Dictator Game. The lower panel shows the choices available for both the FM and SM in the Investment Game.

Hormone measures

Blood OT levels are highly correlated with cerebro-spinal fluid OT levels (Carson et al., 2015); therefore, hormone levels were assessed using blood samples. In order to draw blood quickly and easily during the course of the experiment, a trained, registered nurse placed an IV catheter at the start of the experimental session. This was usually placed in the median cubital, cephalic, or basilic vein of the arm; however, if the nurse encountered any difficulty placing the catheter in the arm, it could be placed in one of the dorsal metacarpal veins.

We drew blood samples at various points in the session to measure OT, ACTH, and cortisol. We took the first sample directly after the catheter was placed, and subsequent blood samples before the instructions to each IG (as a baseline before the interaction with the partner and before any decisions had been made) and then directly after each dyadic interaction prior to the IG (within the 3-minute half-life of blood oxytocin).

At each draw, we took 15mL of blood, cooled it in ice, and treated it with aprotinin to slow degradation. We centrifuged the samples at 1700g for 14 minutes at 3˚C. Plasma was extracted and all samples stored in an ultracold freezer at -80˚C prior to offsite analysis using Bachem RIA (radioimmunoassay) kits. Extraction solvents differed between analysis sites with some using 98% acetone and others 60% acetonitrile. Validation tests revealed greater accuracy for the latter technique (Christensen, Shiyanov, Estepp, & Schlager, 2014 [reprinted in this issue]); therefore, results reported here are based on use of the 60% acetonitrile extraction solvent.

Psychophysiology measures

We measured EEG, ECG, respiration, and EDA at a number of points during the experiment. Recording took place during an eight minute quiet period when the participant was not directly involved in any tasks; participants kept their eyes open for four of the eight minutes and closed for the other four. These quiet periods took place at the beginning of the session following initial instrumentation (baseline), following the DG, following the face-to-face interactions prior to the IG, and after the results of the game randomly chosen to determine participant payments (see Figure 1). We report results for the measures taken following dyadic interaction in this paper.

We recorded EEG at 1000 Hz using 64 channel Geodesic EEG 400 systems and NetStation software (Electrical Geodesics, Inc.; Eugene, OR). ECG and respiration were recorded using additional polygraph inputs to the EEG system. ECG was recorded using a three-lead system, with electrodes placed above the right clavicle, below the left ribs and above the left clavicle. We derived four measures from the ECG recordings: heart rate, and three measures of Heart Rate Variability (HRV) - log high-frequency power, wide proportional high-frequency power, and the ratio of low- to high-frequency power. We measured respiration using bands placed around the thorax (under the armpits) and the abdomen. For EDA data, we used an MP150 data acquisition system and Acqknowledge software (BIOPAC; Goleta, CA). We placed electrodes on the index and middle fingers of the participant's non-dominant hand, and operationalized EDA in two ways: as skin conductance level (SCL), i.e., the mean level during the quiet period; and as non-specific skin-conductance responses (SCR), i.e., the number of positive deflections in the EDA recording greater than 0.1µS.

Survey measures

In addition to the behavioral, hormonal, and electrophysiological measures, we also administered a number of surveys and self-report measures. Participants completed the Social Values Orientation survey, the Emotion Regulation questionnaire, the Big Five personality questionnaire, the Positive and Negative Affect Schedule (PANAS), a Risk-Taking questionnaire, a demographic questionnaire, a Friendship Closeness survey, an Empathy survey, the Stephenson Multigroup Acculturation Scale, and an exit survey. Following their dyadic interactions, participants also indicated their subjective level of trust in their partner on a five-point scale.

Results

Behavioral Results

Figure 3 shows the behavioral results grouped by relationship status (i.e., friend or stranger). The left panel shows the distribution of altruism decisions in the DG (i.e., the proportion of decisions in which each amount of money was transferred). Participants sent $120 (the maximum amount) to their friends 65% of the time whereas they sent $120 to strangers only 14.1% of the time. This suggests that either participants were more altruistic towards their friends than strangers, or (as was confirmed by comments in the exit survey) at least some friend pairs agreed to split the money outside the experiment (essentially turning the DG into an IG). The right panel shows the distribution of trust decisions in the IG. Participants sent $120 to friends 87% of the time and 53.4% of the time to strangers. The lack of within-subject variability in trust behavior (particularly between friends) made it difficult to calculate meaningful correlations between trust decisions and neurophysiological variables in Protocol A. Therefore, to focus on the widest range of trust behavior and to ensure no collusion outside the experiment, we restricted our primary analyses to Session 1 of Protocol B, where all participants were strangers without any relationship history. Behavioral results from Session 1 were most similar to those found in the existing literature, trust in partners was more variable, and participants were most likely to be concerned with the consequences of defecting on their partner since they would see them again the following week. Table 1 shows the behavioral results for this subset of the data. There were 138 decisions where participants transferred a larger amount in the IG than in the DG (i.e., demonstrating trust) and 82 decisions where they transferred the same amount in the IG and DG (i.e., demonstrating lack of trust). In 12 decisions, participants transferred more in the DG than in the IG.

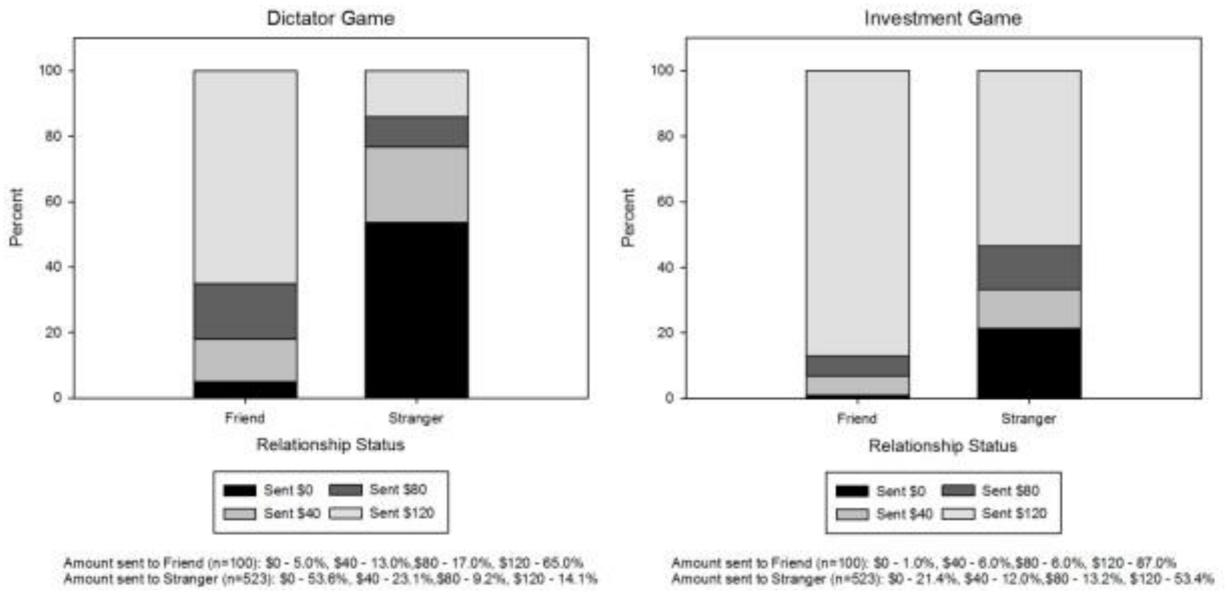


Figure 3. Distribution of amount sent in the DG (left) and as FM in the IG (right) grouped by relationship type, across all protocols and sessions.

Table 1. Distribution of DG and IG decisions during Session 1 of Protocol B (i.e., all participants were strangers). Light shading indicates trust (where participants sent more in the IG than the DG). Medium shading indicates no trust (where participants sent the same amount in both games). Dark shading indicates active distrust (where participants gave less in the IG than the DG).

		Dictator Game Decision				
		0	**40**	**80**	**120**	**Total**
Investment Game Decision	**0**	36	4	1	0	41
	40	16	12	1	2	31
	80	12	16	11	4	43
	120	42	36	16	23	117
	Total	106	68	29	29	232

IG>DG	IG=DG	IG<DG

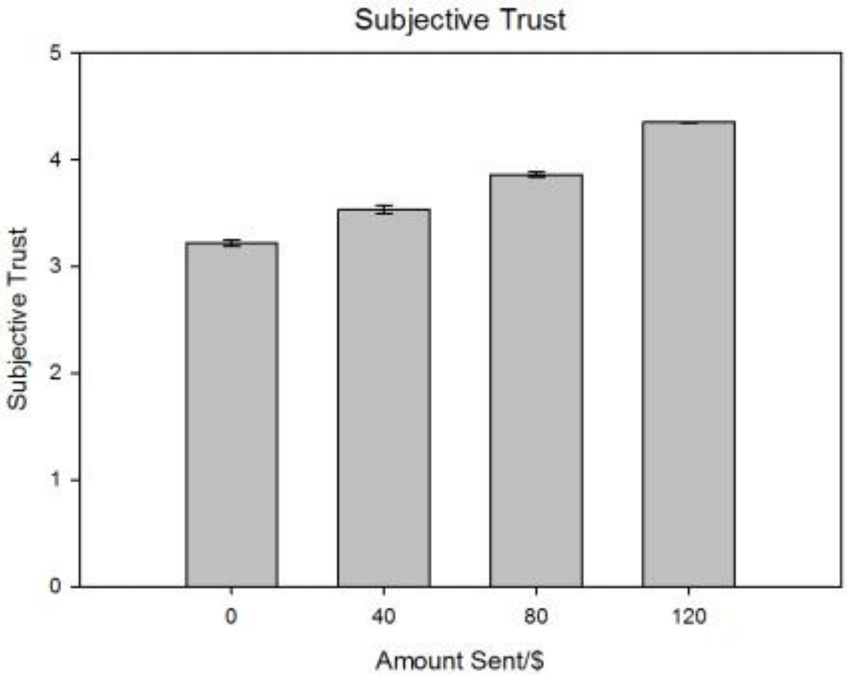
Neurophysiological results

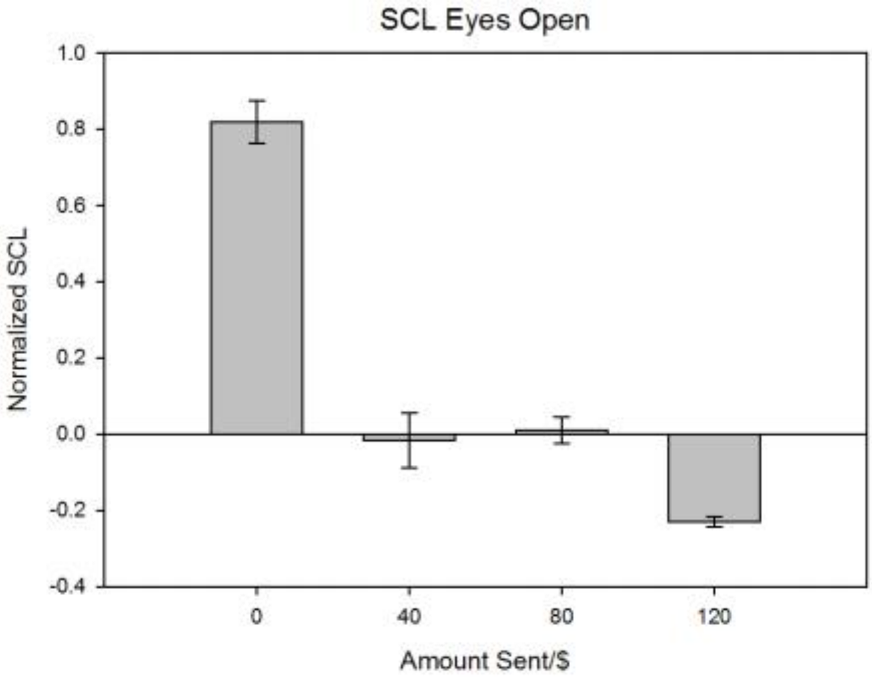
We used three approaches to investigate the relationship between the behavioral, neurophysiological, and subjective measures of trust: a regression analysis of each measure individually on behavioral trust, a regression analysis of all measures together on behavioral trust, and a correlation analysis of the individual participants' neurophysiological measures with trust (regressions were ordinal logistic regressions). We conducted the first two analyses on data from Session 1 of Protocol B, and the third one on all the available data from both sessions of Protocol B. Trust was operationalized as the amount sent by the FM in the IG; we controlled for altruism by entering the amount sent in the DG as a covariate.

In the first set of analyses, we initially ran regressions with behavioral trust as the dependent variable, altruism as a covariate, and individual physiological or subjective trust measurements as independent variables. This analysis generates two values—incremental R^2 and percent improved predictions. Incremental R^2 is a measure of the improvement in the regression model's fit with the inclusion of the independent variable. Percent improved predictions is a measure of how much improvement the model shows in predicting behavioral trust with the addition of the independent variable. (A simulated regression analysis on a dataset of random variables of a similar size showed that we would expect 52% improved predictions by chance.) We found a weak relationship between behavioral and self-reported measures of trust ($p < 0.01$, incremental

$R^2 = 0.07$, percent improved predictions = 77%). We also found a weak relationship between behavioral trust and SCL during quiet periods (eyes open: $p < 0.01$, incremental $R^2 = 0.05$, percent improved predictions = 61%; eyes closed: $p < 0.05$, incremental $R^2 = 0.04$, percent improved predictions = 61%). In the logistic regression analysis with an indicator of distrust as the dependent variable (how much less FM participants sent in the IG than the DG, for rounds where participants gave more in the DG), we found a significant relationship with two measures of HRV. Wide proportional high frequency power (a measure of vagal modulation of heart rate and parasympathetic tone), and the ratio of the low to high frequency power (the balance between sympathetic and parasympathetic activation, or sympatho-vagal balance) during the eyes open quiet period, both related to distrust (wide proportional high-frequency power: $p < 0.01$, incremental $R^2 = 0.13$, percent improved predictions = 73%; low- to high-frequency power ratio: $p < 0.01$, incremental $R^2 = 0.04$, percent improved predictions = 67%).

Subsequently, we ran the first set of regression analyses with both altruism and subjective trust as covariates. The relationships between trust and skin conductance and heart rate variability were maintained. However, we also found two other relationships between hormones and trust. ACTH was associated with both trust ($p < 0.05$, incremental $R^2 = 0.01$, percent improved predictions = 64%) and distrust ($p < 0.05$, incremental $R^2 = 0.05$, percent improved predictions = 80%). OT also showed a negative relationship with distrust ($p < 0.01$, incremental $R^2 = 0.03$, percent improved predictions = 64%). These results suggest that SCL and ACTH measures contribute slightly to predictive accuracy above that available with the subjective measure of trust alone, and that HRV and OT are possibly associated with distrust. Figure 4 shows the relationships with behavioral trust for the three variables (subjective trust, SCL, and ACTH) that showed a significant association.


Subjective Trust


SCL Eyes Open

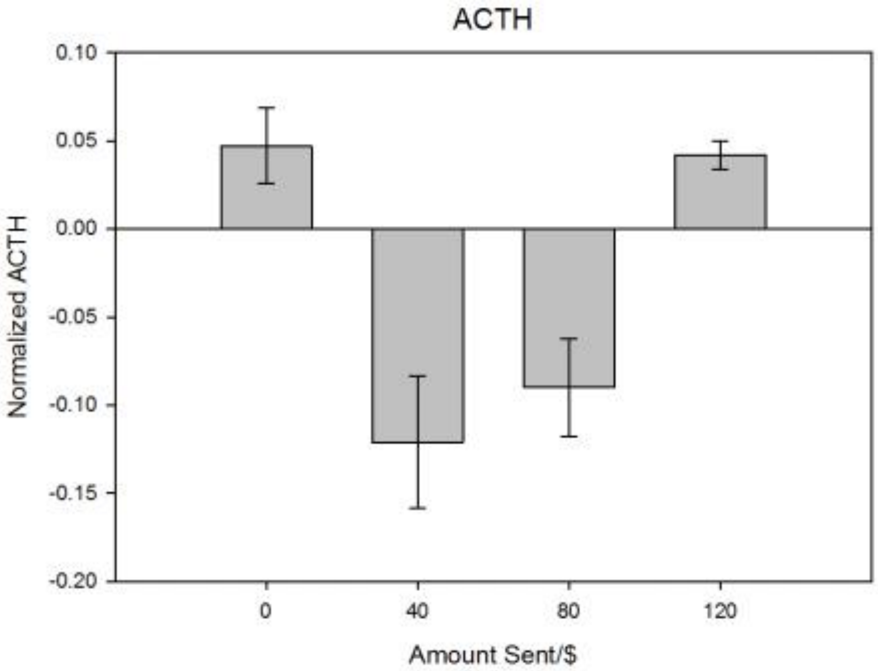


Figure 4. Plots showing the relationship between trust behavior and subjective trust, SCL (during the eyes open quiet period), and ACTH. Error bars represent SEM.

For the second analysis, we ran a regression that included all the neurophysiological variables as independent variables and behavioral trust as the dependent variable. Missing data (e.g., when we were unable to draw blood, or when electrodes came loose) were imputed using two methods—mean value substitution and multiple imputation—both of which yielded similar results. Using the amount sent as FM in the IG as the dependent variable, the collection of all physiological measurements increased R^2 above the amount from just altruism alone by 0.046 ($p = 0.034$). With both altruism and subjective trust entered as covariates, the incremental R^2 obtained by entering the physiological measurements as covariates was 0.053 ($p = 0.0003$). This suggests that the neurophysiological measures better predict the amount sent in the IG than do altruism and subjective trust alone.

We conducted a final set of analyses to determine if trust and physiological measurements were highly correlated for any subset of individuals. We correlated each participant's trust behavior with each physiological variable using all the available data for Protocol B. For each subject, we then averaged the absolute values of these correlations across all physiological variables to create an overall subject-specific correlation. A permutation analysis tested whether the distribution of participant correlations significantly differed from that which would be expected by chance; the results indicated it did not.

Trustworthiness results

In addition to the analyses of trust behavior, we also conducted similar analyses with respect to trustworthiness of both partners. Trustworthiness was defined as how much the participant would return as SM. The results revealed a relationship between a participant's trustworthiness and their subjective rating of trust in their partner ($p < 0.01$, incremental $R^2 = 0.03$, percent improved predictions = 69%). We also found a weak relationship between SCR and trustworthiness during the eyes closed quiet period ($p < 0.05$, incremental $R^2 = 0.01$, percent improved predictions = 60%), which remained when subjective trust was also included as a covariate in the regression.

Discussion

This project sought an extensive analysis of the relationships between trust, and hormonal and physiological measures. Additional goals included assessing the differences in trust between friends and strangers, between American- and foreign-born participants, and over time. Analysis of Protocol A revealed a very high level of trust between friends and evidence for collusion outside the experiment. Consequently, within friendship dyads, the minimal variability in trust did not allow a meaningful analysis of its relationship with neurophysiology. As a result, we developed Protocol B, in which only strangers participated; this resulted in greater variability in trust behavior. In Protocol B, we also recruited both US and foreign born participants; in this way we hoped to be able to test cultural differences in trust, but given the weak relationships we discovered when considering the sample as a whole, we consider this experiment underpowered to test differences across cultural groups. We do, however, believe this would be an interesting direction for future research.

In addition to the behavioral measure of trust (i.e., how much the participant sent as the FM in the IG), we also included a self-report measure. While the results of the regression analysis were in the expected direction—i.e., the more trust people reported, the more money they sent to their partner—the relationship was weak, with only a 7% increase in predictive accuracy when covarying for altruism. There may be several reasons for this weak relationship—including the the different measurements of trust. For example, in the subjective measure of trust, participants may have been subject to social desirability effects and been reluctant to report their lack of trust to the experimenters, or they may have promised to send a certain amount during the earlier interaction, and felt obliged to send that amount despite their lack of trust. For the behavioral measure of trust, participants may have been subject to social norm or social pressure effects, and felt obliged to transfer money even when they didn't trust their partner. Previous research (Glaeser, Laibson, & Scheinkman, 2000) has also shown that behavioral measures of trust don't necessarily have a strong relationship with simple self-report measures of trust.

Our analyses revealed only a few weak relationships between neurophysiological measures and trust. SCL was the only neurophysiological measure associated with trust, and it only led to a 5%

increase in predictive power at most. SCL decreased with increasing trust, perhaps reflecting decreased anxiety that the trust would not be reciprocated.

Only when covarying for both altruism and subjective trust did ACTH emerge as significantly increasing predictive power, and only by 1%. This effect was not monotonic: ACTH increased from baseline when the participant gave either nothing or $120, but decreased when she gave $40 or $80. This result contradicts our prediction that stress (reflected by ACTH) would decrease with trust. One possible interpretation of this result highlights a confound in the experimental design between behavioral trust and risk. When participants send more money in the IG, they both display more trust and assume more risk. ACTH (at least for decisions that involved sending money) might track perceived risk rather than trust. The increase observed when participants sent no money might have a number of explanations: participants might worry that sending nothing is the wrong thing to do since it precludes any increase in payout, they may be reneging on the promise made to their partner or it might reflect violation of a social norm.

While we found only weak relationships between trust and SCL and ACTH, we found no relationship between trust and frontal alpha asymmetry, heart rate, heart rate variability, OT or cortisol. The relationship between distrust, OT, and some measures of HRV is based on analyzing only the few instances when participants transferred more money in the DG than the IG; therefore, these results may not be reliable. Even considering all neurophysiological variables together, the additional predictive power when covarying for altruism and subjective trust was only 5.3%. Given the relationship between trust, OT, and other psychophysiological measures reported in previous studies (Baumgartner et al., 2008; Zak et al., 2004), it was quite surprising to find so few significant associations.

The limited positive findings may accurately reflect the greater variability of trust-related signals in an unscripted interaction compared with a more tightly controlled experimental design. If this is the case, we would expect minimal predictiveness to generalize in the field. However, some of the lack of consistent neurophysiology findings could also be due to limitations of this protocol. For example, we elected to make our primary neurophysiological measurements directly after the interaction, under the assumptions that trust (or distrust) built up during the interaction, and that these emotional responses would be maintained immediately afterward. However, it is possible that these interactions may have resulted in limited trust judgments (e.g., because of the lack of feedback on partner behavior), or that the timing of physiological responses did not overlap with the timing of measurements. One improvement that could be applied to future research might be to make physiological measurements at more points during the experiment or each physiological response to trust could be investigated in a smaller experiment optimized for that particular measure.

Further, during each interaction in our study, participants discussed their strategy as both the trustor and trustee. Therefore, the measures taken following the interaction could be related to processes involved in both assessing their partner’s trustworthiness and planning their own behavior. A cleaner experimental design that could be adopted in the future would have each participant consider strategy as only one of the players.

The positive relationship between a participant's trustworthiness and trust in their partner suggests that trusting people are also trustworthy. Interpreting the positive relationship between SCR and the trustworthiness of the participant's partner is difficult. First, it only occurred during the eyes closed quiet period, not during the eyes open period. Second, if SCR reflects periods of transient emotional arousal, we would expect an increase in partner's trustworthiness to be associated with a decrease in SCR (i.e., a negative relationship).

In summary, we failed to find a relationship between trust, trustworthiness, and many of the neurophysiological variables we recorded, and where we did find relationships, they were relatively weak. However, interest in the neuroscience of trust (and other complex psychological states) is increasing. Improved neuroscience methodologies, experimental paradigms, and analysis techniques offer hope for the future of research into trust that may build on the initial forays reported in this and the accompanying articles.

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via United States Air Force Research Laboratory contract number FA8650-11-C-7113. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

The authors would like to acknowledge the assistance of Adam Guastella and Andrew Kemp during the early development of the experimental protocol and the anonymous reviewers for their insightful and helpful comments.

References

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, *58*(4), 639–650. http://doi.org/10.1016/j.neuron.2008.04.009

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. http://doi.org/10.1006/game.1995.1027

Carson, D. S., Berquist, S. W., Trujillo, T. H., Garner, J. P., Hannah, S. L., Hyde, S. A., … Parker, K. J. (2015). Cerebrospinal fluid and plasma oxytocin concentrations are positively correlated and negatively predict anxiety in children. *Molecular Psychiatry*, *20*(9), 1085–1090. http://doi.org/10.1038/mp.2014.132

Christensen, J. C., Shiyanov, P. A., Estepp, J. R., & Schlager, J. J. (2014). Lack of association between human plasma oxytocin and interpersonal trust in a prisoner's dilemma paradigm. *PLoS ONE*, *9*(12), e116172. http://doi.org/10.1371/journal.pone.0116172

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*(6), 1585–1593. http://doi.org/10.1016/j.jrp.2008.07.011

Glaeser, E. L., Laibson, D. I., & Scheinkman, J. A. (2000). Measuring trust. *Quarterly Journal of Economics*. http://doi.org/10.2307/2586897

Harmon-Jones, E., & Allen, J. J. B. (1998). Anger and frontal brain activity: EEG asymmetry consistent with approach motivation despite negative affective valence. *Journal of Personality and Social Psychology*, *74*(5), 1310–1316. http://doi.org/10.1037//0022-3514.74.5.1310

Heinrichs, M., Baumgartner, T., Kirschbaum, C., & Ehlert, U. (2003). Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biological Psychiatry*, *54*(12), 1389–1398. http://doi.org/10.1016/S0006-3223(03)00465-7

Insel, T. R., Winslow, J. T., Wang, Z., & Young, L. J. (1998). Oxytocin, vasopressin, and the neuroendocrine basis of pair bond formation. *Advances in Experimental Medicine and Biology*, *449*, 215–224.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–676. http://doi.org/10.1038/nature03701

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., … Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20084–20089. http://doi.org/10.2307/25450845?ref=search-gateway:0746ce012b3e8b38c95cce945a0158cd

Lee, H.-J., Macbeth, A. H., Pagani, J., & Young, W. S. (2009). Oxytocin: The great facilitator of life. *Progress in Neurobiology*. http://doi.org/10.1016/j.pneurobio.2009.04.001

Levin, I. P., Xue, G., Weller, J. A., Reimann, M., Lauriola, M., & Bechara, A. (2012). A neuropsychological approach to understanding risk-taking for potential gains and losses. *Frontiers in Neuroscience*, *6,* 15. http://doi.org/10.3389/fnins.2012.00015

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022. http://doi.org/10.1177/0149206306294405

Luo, A. H., Tahsili-Fahadan, P., Wise, R. A., Lupica, C. R., & Aston-Jones, G. (2011). Linking context with reward: A functional circuit from hippocampal CA3 to ventral tegmental area. *Science*, *333*(6040), 353–357. http://doi.org/10.1126/science.1204622

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. http://doi.org/10.2307/258792?ref=search-gateway:f1909b743e505d08717ccd0d51e778a9

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *The Academy of Management Journal*, *38*(1), 24–59. http://doi.org/10.2307/256727?ref=search-gateway:dc82a00bfb8141ea54b51c45cd37ea4b

Mooradian, T., Renzl, B., & Matzler, K. (2006). Who trusts? Personality, trust and knowledge sharing. *Management Learning*, *37*(4), 523–540. http://doi.org/10.1177/1350507606073424

Naef, M., & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.1367375

Powell, E. W., & Rorie, D. K. (1967). Septal projections to nuclei functioning in oxytocin release. *American Journal of Anatomy*, *120*(3), 605–610. http://doi.org/10.1002/aja.1001200310

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*(5850), 598–602. http://doi.org/10.1126/science.1142996

Schilke, O., Reimann, M., & Cook, K. S. (2013). Effect of relationship experience on trust recovery following a breach. *Proceedings of the National Academy of Sciences*, *110*(38), 15236–15241. http://doi.org/10.1073/pnas.1314857110

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). A sniff of trust: Meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology*, *37*(3), 438–443. http://doi.org/10.1016/j.psyneuen.2011.07.008

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283. http://doi.org/10.1038/nn816

Zak, P. J., Kurzban, R., & Matzner, W. T. (2004). The neurobiology of trust. *Annals of the New York Academy of Sciences*, *1032*(1), 224–227. http://doi.org/10.1196/annals.1314.025

Zak, P. J., Kurzban, R., & Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Hormones and Behavior*, *48*(5), 522–527. http://doi.org/10.1016/j.yhbeh.2005.07.009